Wickliffe W. Shreve II. CrowdTrusting: Case Studies in Crowdsourcing Projects. A Master's Paper for the M.S. in L.S degree. November, 2015. 48 pages. Advisor: Denise Anthony

Crowdsourcing has gained popularity over the past few years as a way for library and archive professionals to supplement and enhance the description of their collections. This paper provides case studies of four community archiving projects, focusing on crowdsourcing techniques they used to describe or enlarge their collections. The studies were conducted to determine the kinds of techniques used in community archives, and the potential benefits and barriers they faced in developing and using the techniques. Analysis of the projects indicated that the up-front investment in developing crowdsourcing tools may be prohibitive for community archiving projects. However, the results also indicated that digitization projects were still of value.

Headings:

Crowdsourcing

Digitization of archival materials

Archives -- Social aspects

Archives -- Public relations

Archivists -- Interviews

CROWDTRUSTING: CASE STUDIES IN CROWDSOURCING PROJECTS

by
Wickliffe W. Shreve II

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

November 2015

Approved by

_____

Denise Anthony

Table of Contents

Introduction

The purpose of this study is to assess whether there are common deficiencies archivists/special librarians or librarians in general have found when designing and leading crowdsourcing projects in community archiving; and therefore, whether crowdsourcing as a tool may be suited to community archives generally. These deficiencies may be in the tools chosen, the management of the project, the type of project, etc. However, they may also be related to personal factors, including a difficulty in relinquishing archival control to the crowd. The deficiencies may also be project-type specific. In order to determine whether such deficiencies exist, the focus will be on in-depth case studies covering as many areas as possible, beginning to end.

The term crowdsourcing is a portmanteau of the words "crowd" and "outsourcing." The first use of the term has been attributed historically to Jeff Howe in an article written for Wired magazine in 2006. Howe defines crowdsourcing as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (Howe, 2006). Wikipedia's current definition, drawn from Merriam Webster, remains close to this original description: "Crowdsourcing is the process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, and especially from

an online community, rather than from traditional employees or suppliers." Some definitions focus on the online community as essential to the crowdsourcing task (Brabham, 2012; Brady, 2014). Regardless, the simplest definition of crowdsourcing has remained relatively consistent over time.

Project leaders turn to crowdsourcing for myriad reasons. One of the more prevalent motivations is the completion of simple tasks that can be completed quickly, most often in order to free up more skilled professionals to concentrate on larger, more complex issues (Alonso et al., 2008; McCreadie, 2013). This also includes the division of larger tasks into smaller portions, sometimes called "distributed human intelligence tasking" (Brady, 2014). The problem that arises from drafting what many consider the "unskilled" crowd is whether this potential initial savings results in work that is useful, or whether the review of the work results in actual higher costs. In particular, the work can be of low quality. Perhaps more insidious, though, certain platforms or projects may result in the crowd actually maliciously interfering with the work (Downs et al., 2010; McCreadie, 2013). At least a cursory understanding of the identity of the crowd, therefore, is an essential element in these projects.

Initially, crowdsourcing was conceived as a business solution (as can be deduced from its roots in the term "outsourcing") (Saxton, 2013; Sutherlin, 2013). Early representative crowdsourcing tasks followed this model by attempting to capture the labor of the crowd, examples which include the creation of iStock photo ("de-professionalizing" the

photography industry) and Wikipedia ("de-professionalizing" the reference industry...and cited above) (Sutherlin, 2013).

The increased presence of online collections has allowed for consideration of crowdsourcing as a tool for management of special library collections and archives. In recent years, users of archives have increasingly demanded greater accessibility of historical records and other materials through digitization (Noll, 2013). Digitization is a first step; tasks which would make the documents accessible (e.g., transcription, translation, and the addition of metadata in order to enable searchability) are essential next steps in order to meet the demands of users. Postmodernism's effect on the archivist's role as a professional speaks to these very points, as users are seen as more active participants in the creation of metadata (and through metadata and other interactions actual new meaning) for themselves and other users of the archives (Nesmith, 2002; Vershbow, 2013). However, there may still be professional push-back as archivists may be unwilling to relinquish their traditional role of control, as many information professionals may collapse the concepts of "expertise" and "control" (Wright, 2010).

The idea that users are active participants in the creation of archives, and archival description, is perhaps most relevant in the context of community archives. The focus on community archives has come about partially due to a recognition that certain communities have been "under-voiced and under-represented," particularly in terms of the actual historical materials and primary resources of those communities (Flinn et al.,

2009). It is further based on the recognition that the stories of these communities are an essential part of the general history of where those communities exist (Flinn, 2007). Community archivists often face the opposite problem of those being unwilling to give up control of description of materials, or even materials themselves; they often have to deal with "creators and custodians of these community archives, who often distrust or are at least wary of the intentions of heritage professionals and may wish to prevent the transfer of their papers and social memory to professional, non-community [organizations]" (Flinn, 2007, p. 163). The combination of recognition of the unique expertise of the community and the willingness to relinquish archival control to "non-professionals" may, therefore, position crowdsourcing as a tool suited to curation and metadata creation for community archives.

Regardless, as online accessibility is becoming (if not already) the norm in contemporary society, the increased demand for more digital materials is evident. This is the case even as funding for projects that increase accessibility of historical documents through digitization and transcription is falling. As much as any other institution faced with the financial pressure of steadily shrinking budgets, archival institutions and other special collections must develop processes more effectively and efficiently to ensure a substantial return on investment. The grant-making arm of the National Archives and Records Administration has seen an over 65% decline in funding over the past 5 years for digitization projects, despite the demand for online accessibility (National Historical Publications & Records Commission, 2015).

Crowdsourcing archival and special collections has been viewed as a potential response to these budget constraints as well as the demand for more electronically available product. Tagging documents has allowed for the creation of "folksonomies" that have increased relevance searches. The National Archives' Citizen Archivist Dashboard and the University of Iowa's DIY History website are two examples of archival institutions that use folksonomy (Noll, 2012). Transcription projects, such as the Old Weather project which has volunteers transcribe weather data from ship logs dating to the 18th century, may have broader implications for other communities (in this case climate change research) (Zimmerman, 2012).

The question remains whether crowdsourcing is a viable solution to these issues, particularly in the realm of community archives. The aim of this paper is to add to the research looking at individual projects to determine where crowdsourcing may be successful; but also, to ask the question, are crowdsourcing techniques at all suited to the development of community archives?

Literature Review

General Crowdsourcing Definition

The invention of Gutenberg's printing press allowed for the written word to be "brought...into the lives of everyday people and took books off the chains in libraries" (Herther, 2012). Crowdsourcing has been presented, through the internet, as something which allows for the participation of the "everyday person" in the creation of new information.

As stated above, a general definition of crowdsourcing is that it is a distributed problem solving and or labor model whereby an organization, often through a specific project, leverages the collective work and/or intelligence of an (most often) online community to achieve certain defined goals (Bergvall-Kåreborn, 2014). This definition, however, can tend to create confusion as to what is or is not, or may or may not be "crowdsourcing" (e.g., open source, open innovation, and commons-based peer production) (Bergvall-Kåreborn, 2014). The original definition of crowdsourcing categorically (and perhaps ironically) relates crowdsourcing to outsourcing rather than peer-production and co-creation, suggesting a more significant hierarchical relationship between the crowd and the "crowdsourcer" than projects that are collaboratively egalitarian. The distinction is important: it is less "uncritically celebratory" of crowdsourcing's apparently inherent ability to democratize work (Bergvall-Kåreborn, 2014).

Crowdsourcing Compared to Other Electronic Collaborative Efforts

When documenting crowdsourcing projects, it is essential to define what they are not as much as what they are. Social media as a phenomenon can easily be confused with crowdsourcing. However, platforms such as Facebook, Twitter, Instagram, Pinterest, etc., are not themselves crowdsourcing projects so much as "community forums" or "gathering spaces." Although the platforms may offer opportunities to be used for specific crowdsourcing projects, they represent the separation between platforms and projects.

Crowdsourcing is also often conflated with efforts at determining, from the crowd, what is popular. Classic examples are news and opinion sites such as Reddit and BuzzFeed. BuzzFeed is obsessed with what's viral, or what can be defined as what's important, timely, interesting, funny, and gossip-worthy--essentially, it works as an online stand-in for what used to happen around the water cooler (Archambault, 2013). Although "going viral" is a crowd phenomenon, it has been described more as a "clever blend of curation and crowdsourcing," and not clearly a crowdsourcing project itself (Archambault, 2013).

Collaborative writing applications ("CWAs") are another example of collaborative efforts that are not clearly "crowdsourcing." CWAs are made up of software that allows users to create online content, with permissions for others to also create, edit and supplement. Wikis (on a smaller scale than Wikipedia) are a "classic" example. They can be open or public, which can be edited and seen by anyone; partially public, which can be seen by anyone but require editing permissions; and closed, which are edited by members of an institution and are visible only to members of the institution (Archambault, 2013).

Though they do not encompass all online collaborative projects, crowdsourcing efforts do include myriad projects.

Informal "polls" and CWAs reveal the difference between crowdsourcing projects and collaborative tools. Collaborative tools may be used for crowdsourcing projects because they offer particular affordances ideal for such projects. Wikis (as examples of CWAs) in particular are good tool examples, revealing that the aspects of trust and scale are important factors in determining whether a project is collaborative or crowdsourcing. Wikipedia is crowdsourcing: first, the scale aspect allows for general participation - the crowd is large and unknown or unregulated at the initial stage, as there are low barriers to entry for edits/changes/additions to the work-in-progress; second, the trust aspect allows for unknown individuals to participate - the element of trust is focused on the inability of the participant to undermine the project (i.e., the changes that an individual member of the crowd makes, when incorrect, will not require significant work or time to correct by a managing agent).

Smaller scale wikis demonstrate the "other side of the story." First, the scale is often limited to those who are known or already involved in the project. Second, the participation is considered to be authoritative, and if changes made by a participant are accepted or rejected there is an understanding that such decisions must be supported if called into question.

Broad Base of Projects Using Crowdsourcing

Crowdsourcing has been applied to a multitude of different projects. Some of the earliest projects have been scientific in nature. One such project was Galaxy Zoo, which invited non-experts to assist in the morphological classification of large numbers of galaxies (Patrick Bagget, 2014). The project is a part of a broader scientific crowdsourcing endeavor called the "Zooniverse," dubbed as a general "citizen science project." From the business perspective, an often-cited example of crowdsourcing is Amazon's Mechanical Turk, which assigns various "microwork," or tiny tasks that take little time and pay very low amounts of money, to individuals. Specifically, a "turker" might be assigned the task of transcribing text from business cards for $0.02 per card, tagging images with subject terms for $0.04 per five-image set, or, increasing the complexity of the task, briefing an industry trend report for $22.50 (Bartlett, 2014).

Although this individual task is being performed by an individual himself or herself, it remains part of the overall crowdsourcing project, as such tasks are continually managed. Further, because the tasks are assigned on a relatively random basis, the member of the project requesting the task is not certain which member of the community will work on it (although often for projects assigned through programs like Amazon's Mechanical Turk, for which participants pay and are paid, there are some reputation controls in place) (Mason, 2012). A crowdsourcing project, therefore, does not require that all aspects of it are changed or edited by the crowd or individual members of it, but rather that they may be. Although in the case of Mechanical Turk, this is almost indistinguishable from distributed manual labor (or perhaps mere electronic outsourcing without the benefit of a

clever portmanteau attached to it), when dealing with projects where the crowd has an interest in the results that is separate from financial remuneration, the base of potential projects expands.

Projects have also extended to include those essential to the law enforcement community as well. Enforcement personnel have used crowdsourcing to determine accurate street prices of black market drugs, including cocaine, heroin, and the off-market sale of prescription medication (Dasgupta, 2013). Law enforcement data requires a profound level of trust, as often it may be used as evidence. This trust must be reciprocal, informants rely on law enforcement personnel assertions that information they provide will not be used against them. Perhaps because of its potential for anonymity, crowdsourcing creates an opportunity for such data gathering.

Taking examples from the business and scientific projects, historical archive projects like the Database of the Smokies ("DOTS") have been established with funding from the American Library Association ("ALA") under the Carnegie-Whitney Grant award program with the "intention of creating a searchable citation database, engaging the public with Smoky Mountain research and publication, and increasing the growth of the database by eliciting unpaid volunteers to assist in identifying and finding published items germane to the Great Smoky Mountains" (Patrick Baggett, 2014). Archivists have looked to the crowd for more "basic" transcription projects, including the Australian Newspapers Digitization Program for text enhancement of historic newspapers as well as taking advantage of the interest of the crowd for the "Transcribe Bentham" project (which

also included the creation of new metadata for the transcribed pieces) (Moyle et al., 2011; Patrick Bagget, 2014).

These projects represent a push for library and archival staff to digitize more, but also make those digitized materials available for the user. The initial digitization produces material that is available in the most basic sense, but does not clearly allow other users to find the new material. It is, in essence, "available" but not "findable." The disconnect in these projects, therefore, seems to lie in the fact that they produce product, but that product is not available to those searching for it. So, do crowdsourcing projects actually produce more product in a meaningful manner?

Motivating the Crowd

Part of the issue for a crowdsourcing project is motivating "buy in" and thereafter continuing participation. Where there is some form of payment, the project can rely on the payment to ensure continuing participation by individual users or users generally. Payments on individual assignments in these scenarios, though, are coupled with guaranteed payment. Therefore, motivation for continuing in the project or program is often related to a reputation rating, so that poor performance prevents participants from being assigned future tasks (Mason, 2012). When the projects are unpaid, however, continued performance is often motivated by some form of "gamification" of the event. This may be related to the reputation rating in a paid environment (e.g., a points system for members of the project). It may also be more literally "gamed." For instance, malaria

researchers have used a web-based game to recruit participants to tag and count malaria parasites on digital images of blood smears (Brady, 2014).

Types of Projects

More Product, Less Process

Crowdsourcing has been offered as a solution to the push for "more product, less process." As far back as 2005, 59% of repositories have acknowledged a major problem with their backloads, and 78% have claimed that each year they take on more product than they have the capability to describe, making the problem potentially exponential (Greene, 2005). The need for alternatives to the classic model of description have been, therefore, readily apparent.

The additional push for digitization creates pressure to make digital objects available before they have been properly prepared, in addition to the accession of the original objects. Part of this is the preservation aspect, including the basic principle of "Lots of Copies, Keep Stuff Safe" ("LOCKSS") (Carr, 2013). This can turn into a "numbers game" in which the quantity outweighs the quality of the information - mainly, that there are many images but little data about the images available. Mainly, the concern about making the information available online becomes primary, where the addition of metadata and other clarifications that would allow the information to be findable as they would be with authoritative and detailed metadata becomes secondary, and crowdsourcing becomes a potential "magical" solution (Woodward, 2014).

Creation of Metadata (e.g., Tagging)

Metadata creation has historically been the domain of the information professional. Allowing the crowd the ability to create metadata is a departure from this norm. Allowing untrained contributors to affix subject tags to citations without editorial review could possibly create a chaotic system that would impede effective browsing, a feature determined to be integral to usability (Patrick Bagget, 2014).

However, many argue that tagging may actually create a more valid vocabulary than one imposed by professionals. Still others argue that the crowd itself acts as a top-down control by creating and editing tags as a group. When allowing the crowd to tag objects with metadata, complete top-down control may actually create a barrier for the user, as it may be "a daunting task for contributors to scroll through and select from a list of subject headings and [therefore] they would be unlikely to do so" (Patrick Bagget, 2014). Current practice appears to embrace a hybrid approach where folksonomies are used to augment, not replace, the more traditional taxonomy methods. For instance, professional indexers can use controlled vocabularies to create consistency and interoperability, supplemented by descriptions from users. Where this may fail to create better metadata, these efforts often succeed in creating user engagement (Patrick Bagget, 2014; Short, 2014). Indeed, it may be essential for the archive or library to maintain relevance (Short, 2014).

Despite the expertise of archival and special collections professionals, the material that they handle often benefits from the local knowledge held by the crowd. For example, the

"Historical Photograph Tagging Project" at the University of Alabama encouraged their users to add relevant information to their thousands of digitized images. The tags created by users were not only added to the images themselves, but were also added to the library's search feature (Zastrow, 2014). Users themselves may also have expertise about a collection beyond that of the archivist or librarian (Zastrow, 2014). Depending on the project, it may be necessary to control for repetitive or irrelevant tags during the management process, even when dealing with populations that have local expertise (Brew, 2010).

On the other hand, depending on budget and the accessibility of the institution's platform, crowdsourcing may be utilized without concern for local or expert knowledge of the crowd. The Library of Congress ("LC") started its "Flickr Commons Project" with the goal of identifying and explaining historical photos by tagging them with keywords and identifying names, occupations, and other information about the people in the photos. The main motivation was deep staff budget cuts, and considering the collection itself was significantly broad, the decision was made to use the popular Flickr platform rather than relying on regular users of the LC's own website.

Transcription

Transcription work these days is often tasked to programs, such as optical character recognition ("OCR") and reCAPTCHA. reCAPTCHA, for instance, although it works mainly as a tool for security against automated programs, is also used to digitize books, newspapers and old-time radio shows. It has been used to transcribe hundreds of millions

of items, achieving word accuracies of up to 99% - but only for small tasks, and tasks that must be performed in order for reCAPTCHA to be effective in protecting resources (Munyaradzi, 2014).

However, there are certain transcription tasks that require human involvement. For academic research, keyword searching may be essential. This requires a high degree of intellectual control, as the material found must be complete and accurate as it may be quoted in the research material produced. Although the simplest approach is to perform full-page OCR, the algorithms struggle when presented with old material with strange typefaces, and other faults (Chrons, 2011).

Transcription projects can be both long and short term. Perhaps one of the best known crowdsourcing projects is Project Gutenberg, founded in 1971 and still the largest single and officially the oldest collection of free electronic books. In 2001, the Project began using “distributed proofreaders” to find, scan, and mark up books in the public domain, page by page, and divided each work into individual pages which can be proofread by several volunteers (Zastrow, 2014). Short term projects that have a prescribed end date can also be extended to include other materials, utilizing an already established platform and tool set. For instance, the University of Iowa Libraries opened its Civil War Diaries and Letters Transcription Project to the public to commemorate the Civil War sesquicentennial. In one year, the project was nearly completed and expanded to include other archival materials (Zastrow, 2014).

Resistance from the Professional Community

Archival work has, traditionally, been considered most effective "when it is unobtrusive or largely invisible," and where archivists have "adopted a strategy of self-effacement" (Nesmith, 2002). Part of this is a professional recognition that, while they have the significant power to describe, and therefore make accessible, the records under their charge, there is often disagreement in that description (Nesmith, 2002). At the same time there is still an understanding that the archivist is the expert in control of describing, transcribing, and finding the objects in a collection.

New conversations about the relationship between the archivist and the user reveal, however, that they may need to engage as co-equals with the user as archives begin to implement Web 2.0 technologies. In fact, having an archivist retain a privileged, authoritative position in an online community actually reduces both the archivist's as well as the crowd's ability to create more product with less process (Daines & Nimer, 2009).

Perhaps because of this, crowdsourcing as a solution has taken much longer to be embraced by cultural heritage institutions such as archives and special collections. Because these institutions provide access to original records and, most importantly, unique manuscripts in a multitude of formats, a single mistake in description or transcription may prevent access to the information object. A typical response to crowdsourcing in these institutions has been: "Oh, not that again ... they've been hounding me about doing that at work—a terrible idea!" (Zastrow, 2014).

Why Community Archives?

Community archives focus on the gathering of materials of various types in order to understand history in a more inclusive and diverse manner than, perhaps, has been the focus of more "traditional" archives (Flinn, 2010). A significant issue has been the definition of a community archive as opposed to a traditional archive, as the inclusive nature of the movement means that such archives can be both small and large, and focus on ethnic, religious, geographic, and other diverse groups (Flinn, 2010).

Community archives, perhaps because of the fact that they represent groups that have been historically ignored or even subjugated, may be more likely to embrace a non-traditional, non-custodial view of archival collection, working with new digital technologies to "expose" the community to the general population (Flinn, 2010). Many of these groups consider their histories (and collections based around those histories) to be "hidden," and look to new technologies to uncover them (Flinn, 2010). Such groups may work to make their materials available to a broader audience on a permanent basis, or may focus on a particular project because of available funding or other opportunities (Flinn, 2010).

Whether these projects can be considered crowdsourcing or collaborative, however, remains up for debate. One example, Greenlevel.org, aimed at creating an interactive South African history based off of a wiki (Flinn, 2010). Although many registered to create content, few ended up doing so (Flinn, 2010). Interest appeared to taper off significantly, such that the site no longer remains active.

Therefore, although we may make assumptions that because we have a group where "buy-in" already exists, and which is made up of those who are experts in and have materials essential to the collection we are focused on, is using a digital, crowdsourcing method the best way to create or enhance the collection?

Methods

Overview

The purpose of this research is twofold: 1) to discover what tools and methods community archivists or librarians working on community projects have used in different projects that may be considered crowdsourcing their collections, and 2) to understand which barriers they have encountered in this process. I have chosen a case study method for this purpose, defined as "an empirical inquiry that investigates a contemporary phenomenon within its real life context when the boundaries between phenomenon and the context are not clearly evident and in which the multiple sources of evidence are used" (Yin, 2014). Although the tools and methods section would be readily discovered through review of the literature or a survey of professionals who have been involved in crowdsourcing projects, the deeper-level information sought consists of the barriers/failures of the process as a whole. Further, although a survey might provide more data that would be quantitative in nature and therefore readily generalizable, this study seeks to determine more "impressions" of the work, which is inherently qualitative. The validity of the case study research is not dependent on its generalization (Lubbe, 2003).

Regardless of the fact that it is not generalizable in a quantitative sense, case studies may jointly study a number of different examples to enquire into a specific phenomenon, population, or condition. These analyses may later provide for a deeper discussion of those issues, and provide roads for further inquiry (Denzin, 2005). In many ways, the depth of the inquiry possible through the case study method is greater than any other

research method, such as a survey (Lubbe, 2003). Further considering that this study was meant to likely outline the potential for future research, generalizability becomes a less significant concern.

Data Collection

Initial contact of participants began via two main fronts: 1) research on crowdsourcing projects that were focused on smaller populations or collections, and 2) personal contacts (e.g., professors, professional contacts, etc.). Both strategies were targeted at the appropriate population: archival and special collection professionals who have led or worked on crowdsourcing projects. The "personal contacts" front yielded the most contacts, as they tended to be more likely to provide the deep information and agree to the time commitment for a relevant case study. Further, these contacts allowed for a "snowball sampling" (i.e., contacts that led to other contacts).

Through these initial contacts, I was able to include four projects in the study. The contacts were selected in an attempt to include a variety of project examples: a personal archive representing a community; a pilot project to begin a community collection; a small, targeted project involving an already established community; and an academic community. Further, selection was based on a variety of institutional examples: private/grassroots; government; and academic.

I initially contacted participants via e-mail, with follow ups made by telephone as necessary, relating the purposes of the study, the expected time commitment, and an

initial outline of the information requested. After initial contact and acceptance from the participants, I conducted semi-structured interviews consisting of basic questions with follow-ups dependent on the responses. Interviews lasted between 30 minutes and an hour, and were conducted both in person and over the phone based on the requests of the interviewee and what was most feasible given the time frame. Interviewees were informed that the conversation was recorded, and that at any point they could request the interview be terminated.

There are multiple concerns, mostly from a validity perspective, about semi-structured interviews, including (but not limited to): (1) the process of collecting data (e.g., the interview process and how it might affect the data collected); (2) internal validity and making sense of data (e.g., whether the interviewee can be considered truthful, either in terms of purposeful withholding or barriers involved with memory); and (3) external validity of the data and findings (i.e., the generalizability) (Diefenbach, 2008).

Variables related to semi-structured interviews, however, are subject to certain controls. These include: (1) the establishment of ethical guidelines; (2) the development of an interview protocol (including introduction and an outline of questions); (3) selection of recording methods and equipment; (4) methods of analysis and collection (discussed below) (Rabionet, 2011).

From an ethical standpoint, in order to maintain confidentiality, transcripts were kept on an external drive which was password protected. Transcriptions, as written, were

scrubbed of personally identifying information, and were kept only for data analysis and drafting of the final paper.

Data Analysis

Interview transcripts were reviewed by hand in order to extract various themes, with particular focus on similar barriers and problems encountered during the project. The extraction of theme was based on common factors discussed in the literature review, as well as new information from the interviews themselves (and data collected). Even though themes were partially based on issues discussed in the literature review, I attempted to avoid assumptions about the types of barriers that may or may not arise during the project. Therefore, encoding of the themes was based on an emergent coding strategy (Williams, 2015). This method of coding is supported by the fact that expectations about the type of data collected (e.g., whether the projects will be transcription, large or small scale, or overall "successful") cannot be known at the outset, and emergent designs generally allow for processes that are not predetermined (Suter, 2011).

Personal bias is a significant concern with this type of methodology. Although it may not come into play in the current study, it cannot be denied that as a researcher I enjoy certain privilege: "[S]ince, in fact, [scientists] have been predominantly university-trained white males from privileged social backgrounds, the bias has been narrow and the product often reveals more about the investigator than about the subject being researched" (Fried,

1979). Further, as someone who lauds crowdsourcing as a solution, my specific bias may play into my analysis of the data gathered.

Results

Organizations and Projects

Project 1

Project 1 involved the personal collection, located in California, of an individual dedicated to telling nearly every aspect of the African-American experience in the United States: family life, participation in arts, entertainment, science, religion, the law, and more. Part of the goal was to ensure that children, all children, were aware of the accomplishments of black people in American history. Over this individual's life, they collected literature, manuscripts, film and ephemera independently. The work was done over 40 years, and resulted in a collection that totaled over one million items.

Most of the 206437was done informally: discovering rare books in garage sales or used bookstores, purchasing photo collections from newspapers and below value, private agreements with holders of film collections, etc. As significant as the collection itself was, there was little work done to preserve or properly house it.

The collection was initially housed in the individual's garage. Open to the public, it grew to include portions of their house. Family members campaigned to find a permanent home for the collection, which was established approximately ten years prior to this writing. Funding for the move came from a variety of sources, individual, corporate, and institutional, pushed by the individual's family and the surrounding community which recognized the value of the collection. Actual space was made available through permission and use of a decommissioned municipal space.

The collection now is maintained almost solely by volunteers with a single executive director, and therefore is not set up to have a dedicated archival professional. Work on Project 1 was made possible through external grants for the particular project, and was limited to two years.

As part of the work to organize and describe the collection, metadata and descriptive cataloging information was entered through proprietary database software, the use of which was provided by the company at a nominal cost. Access to this software was made possible because of campaigning by the family, describing the type of work that would be necessary in order to create a research-quality collection. The majority of the collection was books, so a good deal of the work was copy cataloging from WorldCat.

Project 1 also included adding metadata to a large collection of photographs of African-American politicians and celebrities from the 1940s and 1950s. This ended up being a less systematic part of the work, but was rather weaved into the overall project of describing and arranging the collection as a whole. The volunteers used were trained as part of the overall project, and the training process involved one full weekend training session discussing overall basics of description, consisting of 12 full hours. During training, while organizing, and while describing the collection, volunteers were supervised.

Project 2

Project 2 is a pilot project, under development, covering the documentary histories of five of America's most historic African-American towns, in an attempt to assume active stewardship and understand, interpret and appreciate these historic places through the lenses of their inhabitants. It includes towns in Alabama, Florida, Louisiana and Mississippi.

The staff and equipment required for the project will be provided by a major southeastern university, in an attempt to shift focus from facilitating the accession of collections into the university's overall collection to reaching out to independent collections and providing support. Part of the goal is to use the history of the towns to promote cultural tourism. By digitizing the documents and other materials of interest, they will be made available to both to town residents and the greater research community, allowing for them to be both preserved and shared.

Of specific focus is one town founded by former slaves in the late 1800s. Two graduate students were available for work during a summer session to reside in the town and do cemetery research, supplementing research there with information from Ancestry.com, oral histories from the town, and obituaries from homes in the area to get information about early town settlers. That information was compiled into a website, created by two additional computer science students available for the project, which allowed for town members, high school students, and others from the area to add or edit information that was posted about these early settlers. The site itself is currently being rolled out to the

town. The hope is that information that people in the town find in their basements and attics will create a better understanding of the genealogy of the town.

Project 3

Project 3 took place in a public library in a university-area town. The library itself has a high circulation rate in comparison with other libraries in the state and a close relationship with the town. It has recently moved into a new space. As part of that move, the library worked in partnership with an out-of-state artist to create a permanent, multi-media installation meant to represent the history of the town and the community as a whole.

The project was funded directly through the town, which allocates 1% of selected capital project expenditures for public art. Solicitations were made to the public through print advertisements in the library, e-mail newsletters produced through an in-house subscription to LibraryAware, the library website, and social media accounts (primarily Twitter and Facebook accounts).

The primary purpose of the project was to uncover what was referred to as the hidden history of the town. People were encouraged to take digital images of not only documents, or submit photographs, but also include images of objects not traditionally included in archives (e.g., letterman jackets, signed footballs, etc.). However, all items were digital images/copies of the materials; people were not asked to donate any original materials. People were also encouraged to create new pieces; for example, to bring in

their favorite books for scanning, or even to take pictures of themselves as they participated in the project.

Several methods were used to collect the digital materials/create the digital copies. Multiple events were held in the library space where people could bring materials for library staff to scan. A website was also created in order to directly submit already-digitized materials on Tumblr, with a Wufoo submission form for the actual items. Finally, scheduled appointments were available in the digital media lab in the library where they could work with a staff member to scan items for collection.

Project 4

Project 4 took place within a private university. It involved the creation and development of a password-protected photo archive accessible only by the university community. Access to the archive requires a university-unique online identification. The archive is made open during alumni events. Originally, it was only made accessible to the staff for "tagging events"; however, once it was made available online, they had consistent access to the materials.

A few years ago, different departments within the university asked whether the Project 4 department wanted their various collections of photographs. When high-level assessment of the collection was done, it was discovered that the department had taken custody of over 13,000 photographs. Although there was institutional knowledge of what should be done (e.g., put them in individual tissue, describe them in a finding aid, etc.), the individual who possessed that knowledge was split between several different

departmental responsibilities, and would not be able to dedicate the necessary time to the work.

The solution was to scan the photographs and then make the digital versions available during alumni events to both (1) tag the items and (2) create buzz around the project. Student funds were available, and so students were hired to scan the photographs and create low-level or initial descriptions for the photographs. A series of students were hired for three years to scan and describe them.

Initially, iPhoto was used on a dedicated terminal, which required that terminals be moved out to the event. This system allowed for a certain level of control; because this institution was particularly concerned with reputation management, they could ensure that information could be approved before it was uploaded to the internet. Eventually, the project moved to Drupal, an open-sourced content management system available on the web, and included controls to ensure that edits and comments could be included as "suggested" before any information could be permanently changed.

Issues

Definition of Crowdsourcing

One of the more interesting results of the interviews was how people defined crowdsourcing. Those involved in Projects 1, 2, and 4 all included analog solutions as part of their definitions. Both Projects 1 and 2 specifically referenced the possibility of including volunteers in-space both arranging and describing the materials. Project 4 stated that crowdsourcing could happen "physically, on paper, or electronically."

Although Project 3 did not mention physical projects when asked the definition of crowdsourcing, Project 3's methods of gathering materials utilized both events in-space (i.e., inviting the community to the library to digitize materials and describe those materials in person), and opening up on-site equipment to upload or digitize materials. Neither of these processes is part of the classic crowdsourcing model, which is meant to enable participants to add metadata, provide materials, or do other tasks remotely, rather than on site.

Projects 1, 2 and 4 also focused on "descriptive cataloging" as the main goal of crowdsourcing. Project 1 stated that "when I think about crowdsourcing, I think about description." Project 2 stated "I think about a well defined project with simple instructions...we need these ten people's records to be filled out." Project 4 stated "in terms of metadata, a group of people contributing to the description of something - a piece or group of pieces."

Project 3 focused on what seem to be the marketing sides of crowdsourcing. The first aspect they focused on was bringing the public's attention to the project, and asking for money, ideas, or other contributions. When asked further about whether they would consider "crowdfunding," the process of seeking donations from a large number of the public through a platform such as Kickstarter, as part of the crowdsourcing process, they stated that it was most often the most important part.

Buy In

Automatic buy-in by participants adding descriptive metadata, or providing materials for digitization, was one of the main benefits of crowdsourcing discussed during the interviews. Project 1 described the volunteers as mainly older people who had more time to dedicate to volunteering, but more importantly knew the collector before their death. Project 1's volunteers also recognized the importance of the collection, as there were almost no major African American collections like it west of the Mississippi River. Project 4 discussed the sense of community created during the projects, describing the reaction of many people participating in tagging and tagging events as "look what they're doing for me." More specifically, they mentioned an instance at one of the events:

> There's this great shot from the 1970s...from a...Moot Court competition...and it's one of those things when this guy's wife finds this picture of her husband and she's like "I remember that suit! I remember that...that's when you won this...competition...and honey honey come and see that and there's so-and-so and bringing more and more people over to the computer to remember...and during alumni weekend that's exactly what they're supposed to be doing.

Project 4 attributed some of the buy-in to an understanding of the participants that working on the project helps to get their institution recognized nationally. Project 2 mentions that there may be some work to get buy-in for the project using this rationale: that the project will help gather materials to ensure that the historic towns are recognized as such and put on the national registry.

Most notable was the fact that there was no mention of any gamification strategies. Project 4 mentioned that there was discussion and concern over how they were going to

get people to participate in the tagging events; however, once the terminals were set up they found that participation was automatic and voluntary.

Funding/Staffing

All projects had significant access to funds, funds for staffing, or were able to leverage available staff and facilities to make the projects work. Project 1 was possible only through an IMLS grant which provided paid training time for the managing archival professional prior to relocation to California, but also a salary significant enough to attract the talent necessary for the project. In addition, conference fees were covered by the grant, so those professionals funded by the IMLS grant in Project 1 were able to reap significant professional exposure from the project.

Project 2 has been able to leverage significant university resources. The pilot project is conceived of as a potential alternative to traditional ingest of personal or community collections, and provides professional counseling to community archives so that they can build and maintain their collections on site. Therefore, the salary for the supervisory archivist has been completely covered by the university. Further, student workers were available over the summer, to dedicate their time solely to the project. Most significantly, two computer science majors were associated with the project, and were able to build a website from the ground up to house all of the genealogical data gathered during the project.

Project 3 was funded through the local government set-aside of 1% of capital projects. In essence, it was money that could be used only for a display project. That the project chosen utilized crowdsourcing techniques to collect materials was, perhaps, a "happy accident." Further, the library was equipped with a full media lab, allowing for conversion of multiple formats (VHS, slides, paper materials), along with a dedicated staff member already in place for converting materials.

Project 4, like Project 2, was able to utilize university personnel and, like Project 3, in-house digitization infrastructure. Funding was available for student workers over a period of three years to scan and provide basic description for the 13,000 photographs the department received. More significantly, the director of the project was in a position that had been created specifically for projects such as this: institutional history projects involving new technologies...or as described, anything that could be considered "bridgey."

Top-Down Problems

As much as there is initial buy-in by the volunteers and there has been funding and staffing made available for these projects, one of the common themes that emerged from the interviews were issues originating from the governing structure. Project 1 had issues with a Board of Directors and an Executive Director that had conflicts with deciding on a vision for the collection: will it be a research library or a community center? Due to that, to some degree, there were issues with fundraising.

Project 2 described the difference between the bottom-up support from the community, which was characterized by complete support for nearly any idea to digitize, collect, and describe the information; and although there was support from the university funding the project and providing the support staff for it, there was still a sense that the university was holding onto an older model of trying to increase the collection of the university itself. Essentially, there was a question of "if it isn't broke, don't fix it."

Tools

The majority of the project made use of tools that are freely (or generally available to all); where special tools were used, they were either donated or created based on leveraging the staff and funding discussed above.

For Project 1, the majority of the work was done using Microsoft Office (specifically Excel and Word). Google Drive was used because they did not have a proprietary server. Project 2, in creating their website, utilized GitHub code available on the internet. They also used a university Archive-It account to take a picture of the website they had developed. As Project 2 is a pilot project, they are unsure whether the data on the website will be continued; therefore, the Archive-It picture of the website will remain regardless of whether the website itself goes dark. For Project 3, the library advertised the project and attempted to solicit materials through an already-existing website, Facebook and Twitter accounts, and email newsletter announcements via LibraryAware. A separate Tumblr website was created with a Wufoo submission form for self-delivered digital materials. Project 4 started using a dedicated computer (already available to the

department) with iPhoto to tag the collection, and later moved to Drupal, accessible via the university's already-present Net ID system.

Project 2, although they had access to university systems, discussed the limitations of use of such systems. When working with the computer science students to create a website that would be designed to sit on university servers, the technology contacts informed them "we can help you cook it, but we're not really sure it's [the university server] going to eat it." In other words, they could suggest how the design would be compatible with the current system, but could not guarantee that it would work with the system when finished, or that it would continue to work after initial testing. In order to ensure that the site would have a permanent seat on the university server, attempts to get permission from "the highest places" was necessary "before [they] could say it's going to go forward."

Project 4 used experience with the iPhoto system to determine design issues for Drupal. A specific case illuminated the issue for Project 4: a photo was posted of a faculty member of whose identity the members of Project 4 were certain. When they allowed edits to the photo's description, the identity of the faculty member was repeatedly edited based on mistakes or assumptions about the person's identity because of their hairstyle at the time. Based on this, they made the decision to lock out general editing permission and create a separate section of the Drupal form to allow for "suggested edits."

Barriers to Digitization

Several potential barriers to digitization were mentioned during the interviews. Project 1 relied on analog tagging parties with volunteers already working with the collection. Part of this was due to the fact that the vast majority of the volunteers were older, and less familiar with the technology necessary to do the tagging online.

However, Project 2 introduced a potentially unconsidered issue. While working on digitization projects of any sort, there may tend to be the assumption that any high school kid knows how to use the internet, or a tablet, or a smartphone. One of the community contacts in Project 2 warned that they should not be surprised if younger students or participants "might not know anything that you're talking about." They might, for instance, have minimal familiarity with search engines, or have to be taught how to navigate websites. This potentially creates a "double doozy" scenario, as success of these community digitization projects sometimes depends on relying on an inter-generational communication between community members, the older of which have the materials and knowledge and the younger of which have a general interest in the project and a corresponding tech savviness.

Further, in remote or under-served areas, wifi or even internet connectivity is not a given. Project 2 is working under the assumption that there will actually be no local connectivity, and therefore will be bringing/creating wifi hotspots to/in the area to assist with digitization and uploading.

Conclusion

Although projects discussed in this paper recognize that community members and volunteers are perhaps the ideal type of group to be motivated by a crowdsourcing projects, all projects had significant barriers to creation of such projects from scratch. Projects 3 and 4 relied on institutional resources (both infrastructure and staff), and had funding sources already in place to draw from for the projects. Project 1 relied on temporary grant funding for professional oversight, and had the cataloging content management system significantly discounted. Project 2 benefitted from access to at least five researchers/project members from the partner university at any one time at no cost at all, and at least temporary access to university servers to house the end product.

Beyond staffing and project management, Project 2 demonstrated the problems with maintaining the product after the project end. The built in assumptions of the project were that it was very likely it would end with the semester, and that the website created for intake would go dark at that point. The goal of taking a "snapshot" of the site using Archive-It would merely capture the data already collected and collated; there would be no crowdsourcing capability thereafter.

Further, the question of whether there is appropriate technology access is central to the issue. Where use knowledge was low, as potentially in Project 2, significant up-front time would be spent on training volunteers and participants on the very basics, let alone using the crowdsourcing product developed. Further, where infrastructure access is limited, it is likely that participation or use of the product would be low.

As stated during the Project 2 interview, “sometimes digitization isn’t the silver bullet that [we] think it is.” For community archives, where often the interest lies in telling the story, “why are we focused on the digital? Isn’t there an analog way?” In essence, when the community is available locally to help describe the material, there is perhaps no need for crowdsourcing tools to reach out to them digitally. And if the product may not be used by the population to add metadata (or otherwise add to or describe the materials) anyway, we run up against the basic question of authenticity; if you’re not getting people involved in the website, you’re likely not telling the right stories.

Due to the potential high investment costs overall of creating a crowdsourcing product, and the availability of highly-motivated community members who are willing to work in an analog environment (and may be more comfortable in such an environment), crowdsourcing may indeed fail to be the “silver bullet” we might hope that it is for community archives. That does not, however, undermine digital preservation solutions, however. In fact, each project recognized that making their collections available digitally in some way was essential to bringing attention to the communities themselves. In the case of Project 4, for instance, the importance of creating and maintaining a digital presence that was about the institution and its members was considered central to recruiting volunteers. In the case of Project 2, getting materials out of people’s houses and into the digital environment was one of the only ways to gain the national attention necessary to ensure inclusion in the national register, and the benefits that go with it.

In the end, the issue of how to describe the material for community archives, and whether crowdsourcing is the most efficient or best way to do so, seems significantly less important than that something gets done. Project 4 members stated this idea rather succinctly: “we’re willing to let go of some of the best practices [for now] in order to ensure that materials are digitized…sometimes you have to do something, you can’t just sit on a pile of stuff.” And perhaps even more succinctly: “Doing nothing is the worst.”

Bibliography

Archambault, Patrick M., et al. "Wikis and Collaborative Writing Applications in Health Care: A Scoping Review." *Journal of Medical Internet Research* 15.10 (2013).

Alonso, Omar, and Ricardo Baeza-Yates. "Design and Implementation of Relevance Assessments Using Crowdsourcing." *Advances in Information Retrieval*. Springer Berlin, Heidelberg (2011): 153-164.

Bartlett, Jennifer A. "Internet Reviews: Crowdsourcing in Libraries and Archives." *Kentucky Libraries* 78.2 (2014): 6-8.

Bergvall-Kåreborn, Birgitta, and Debra Howcroft. "Amazon Mechanical Turk and the Commodification of Labour." *New Technology, Work and Employment* 29.3 (2014): 213-223.

Brabham, Daren C. "The Myth of Amateur Crowds: A Critical Discourse Analysis of Crowdsourcing Coverage." *Information, Communication & Society* 15.3 (2012): 394-410.

Brady, Christopher J., et al. "Rapid Grading of Fundus Photographs for Diabetic Retinopathy Using Crowdsourcing." *Journal of Medical Internet Research* 16.10 (2014).

Brew, Anthony, Derek Greene, and Pádraig Cunningham. "Using Crowdsourcing and Active Learning to Track Sentiment in Online Media." *ECAI*. 2010.

Carr, Melissa Eleftherion. "Crowdsourcing Content to Promote Community and Collection Development in Public Libraries." *Journal of Electronic Resources Librarianship* 25.4 (2013): 313-316

Chrons, Otto, and Sami Sundell. "Digitalkoot: Making Old Archives Accessible Using Crowdsourcing." *Human Computation*. 2011.

"Crowdsourcing." *Wikipedia: The Free Encyclopedia.* Wikimedia Foundation, Inc. Web. 2 Nov. 2015. <https://en.wikipedia.org/wiki/Crowdsourcing>

"Crowdsourcing Project Helps Finnish Library to Digitise Historical Documents." IM@T.Online. Feb/Mar (2011): 16-16.

"Crowdfunding." *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. Web. 2 Nov. 2015. <https://en.wikipedia.org/wiki/Crowdfunding>

Daines, J. G., and Cory L. Nimer. "Web 2.0 and Archives." *The Interactive Archivist: Case Studies in Utilizing Web 2.0 to Improve the Archival Experience* (2009).

Dasgupta, Nabarun, et al. "Crowdsourcing Black Market Prices for Prescription Opioids." *Journal of Medical Internet Research* 15.8 (2013): 1-16.

Diefenbach, Thomas. "Are Case Studies More than Sophisticated Storytelling?: Methodological Problems of Qualitative Empirical Research Mainly Based on Semi-structured Interviews." *Quality & Quantity* 43.6 (2009): 875-894.

Denzin, Norman K. The SAGE Handbook of Qualitative Research. 3rd ed. Thousand Oaks: Sage Publications, 2005. Print.

Downs, Julie S., et al. "Are Your Participants Gaming the System?: Screening Mechanical Turk Workers." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010.

Flinn, Andrew. "Community Histories, Community Archives: Some Opportunities and Challenges 1." *Journal of the Society of Archivists* 28.2 (2007): 151-176.

Flinn, Andrew. "Independent Community Archives and Community-Generated Content 'Writing, Saving and Sharing our Histories'." *Convergence: The International Journal of Research into New Media Technologies* 16.1 (2010): 39-51.

Flinn, Andrew, Mary Stevens, and Elizabeth Shepherd. "Whose memories, whose archives? Independent community archives, autonomy and the mainstream." *Archival Science* 9.1-2 (2009): 71-86.

Fried, Barbara. Women Look at Biology Looking at Women : a Collection of Feminist Critiques. Boston: G. K. Hall, 1979. Print.

Greene, Mark A., and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68.2 (2005): 208-263.

Herther, Nancy K. "From Book Reviewing to Crowdsourcing: The Radical Shift in Connecting People to Content." *Searcher* 20.5 (2012).

Howe, Jeff. "The Rise of Crowdsourcing." *Wired Magazine* 14.6 (2006): 1-4.

Lubbe, S. "Development of a Case Study Methodology in the Information Technology (IT) Field in South Africa: A Step-by-Step Approach." *South African Journal of Information Management* 5.4 (2003).

Mason, Winter, and Siddharth Suri. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44.1 (2012): 1-23.

McCreadie, Richard, Craig Macdonald, and Iadh Ounis. "Identifying Top News Using Crowdsourcing." *Information Retrieval* 16.2 (2013): 179-209.

Moyle, Martin, Justin Tonra, and Valerie Wallace. "Manuscript Transcription by Crowdsourcing: Transcribe Bentham." *Liber Quarterly* 20.3/4 (2011): 347-356.

Munyaradzi, Ngoni, and Hussein Suleman. "A System for High Quality Crowdsourced Indigenous Language Transcription." *International Journal on Digital Libraries* 14.3-4 (2014): 117-125.

National Historical Publications & Records Commission. Web. 15 Mar. 2015. <http://www.nhalliance.org/advocacy/funding-priorities/national-historical-publications-records-commissio-2.shtml>

Nesmith, Tom. "Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives." *American Archivist* 65.1 (2002): 24-41.

Noll, Aaron G. "Crowdsourcing Transcriptions of Archival Materials." (2013).

O'Leary, Mick. "BuzzFeed and the New Virality." *Information Today* April (2014): 20-21.

Patrick Baggett, Mark, et al. "Populating the Wilderness: Crowdsourcing Database of the Smokies." *Library Hi Tech* 32.2 (2014): 249-259.

Rabionet, Silvia E. "How I Learned to Design and Conduct Semi-Structured Interviews: An Ongoing and Continuous Journey." *Qualitative Report* 16.2 (2011): 563-566.

Saxton, Gregory D., Onook Oh, and Rajiv Kishore. "Rules of Crowdsourcing: Models, Issues, and Systems of Control." *Information Systems Management* 30.1 (2013): 2-20.

Saylor, Nicole, and Jen Wolfe. "Experimenting with Strategies for Crowdsourcing Manuscript Transcription." *Research Library Issues, December* (2011): 9-14.

Short, Jessica. "Take Ten to Tag! Smithsonian Gardens Public Tagging Initiative." *Technical Services Quarterly* 31.4 (2014): 319-331.

Suter, W. Newton. Introduction to Educational Research: A Critical Thinking Approach. Sage, 2011.

Sutherlin, Gwyneth. "A Voice in the Crowd: Broader Implications for Crowdsourcing Translation During Crisis." *Journal of Information Science* 39.3 (2013): 397–409.

Vershbow, Ben. "NYPL Labs: Hacking the Library." *Journal of Library Administration* 53.1 (2013): 79-96.

Williams, J. Patrick. "Emergent Themes." The Sage Encyclopedia of Qualitative Research Methods. Ed. Lisa M. Given. Thousand Oaks, CA: SAGE Publications, Inc., 2008. 249-50. SAGE Research Methods. Web. 28 Mar. 2015.

Woodward, Eddie. "Metadata for Image Collections: Inverse Proportions: The Quantity vs. Quality Conundrum." *American Libraries Magazine,* June (2014): 42-44.

Wright, A. "Online, it's the Mouse that Runs the Museum." *New York Times. New York Times,* 19 Jan. 2010. Web. 15 Mar. 2015. Available at: http://www.nytimes.com/2010/01/20/arts/design/20museum.html?pagewanted=all&_r=0

Yin, Robert K. Case Study Research : Design and Methods. Fifth ed. Los Angeles: SAGE, 2014. Print.

Zastrow, Jan. "Crowdsourcing Cultural Heritage: 'Citizen Archivists' for the Future." *Computers in Libraries* 34.8 (2014): 21-23.

Zastrow, Jan. "The Opinion Piece: Digital Changes Everything: The Intersection of Libraries and Archives." *Computers in Libraries* 33.9 (2013): 16-18.

Zimmerman, Stephanie & Portz, Hallie. "Historical Data Recovery Through Crowdsourcing" *DttP: Documents to the People* 40.4 (2012): 16-20.

"Zooniverse (citizen science project)" Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., 15 Mar. 2015. Web. 15 Mar. 2015. <http://en.wikipedia.org/wiki/Zooniverse_(citizen_science_project)>